

Sampling

- What is it?
- Why do it?
- How is it done?

Definitions

- A sample is a selected subset of the population that is used to gather information about the population
- Sampling is the process of selecting a number of units from the population

- The term population (or universe) refers to the whole collection of units or elements (Persons, Records, or Events)

Benefits of Sampling

1. Reduction in time
2. Reduction in labour
3. Reduction in cost
4. Improved quality of data

Provided the sample is representative

If we take a large sample will the sample mean be exactly equal to the population mean?

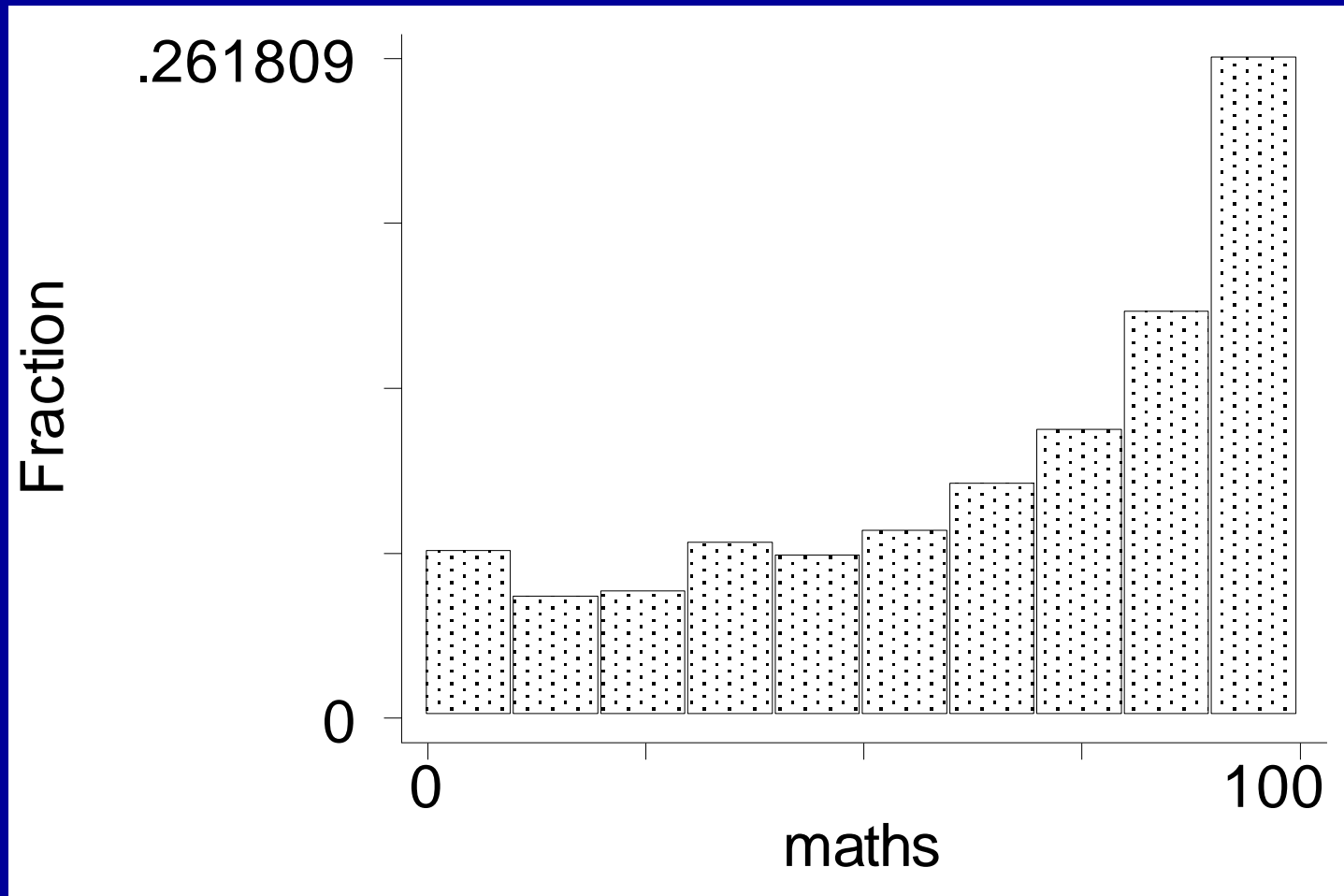
If we take two large samples of equal size will the two sample means be exactly the same?

Health and school performance data

- 2731 Grade 4 children
- Mean (sd) mathematics marks – 64 (29.6)
- Minimum- 00
- Maximum – 100

- Results of repeated samples -

Histogram of mathematics marks of 2731 grade 4 children

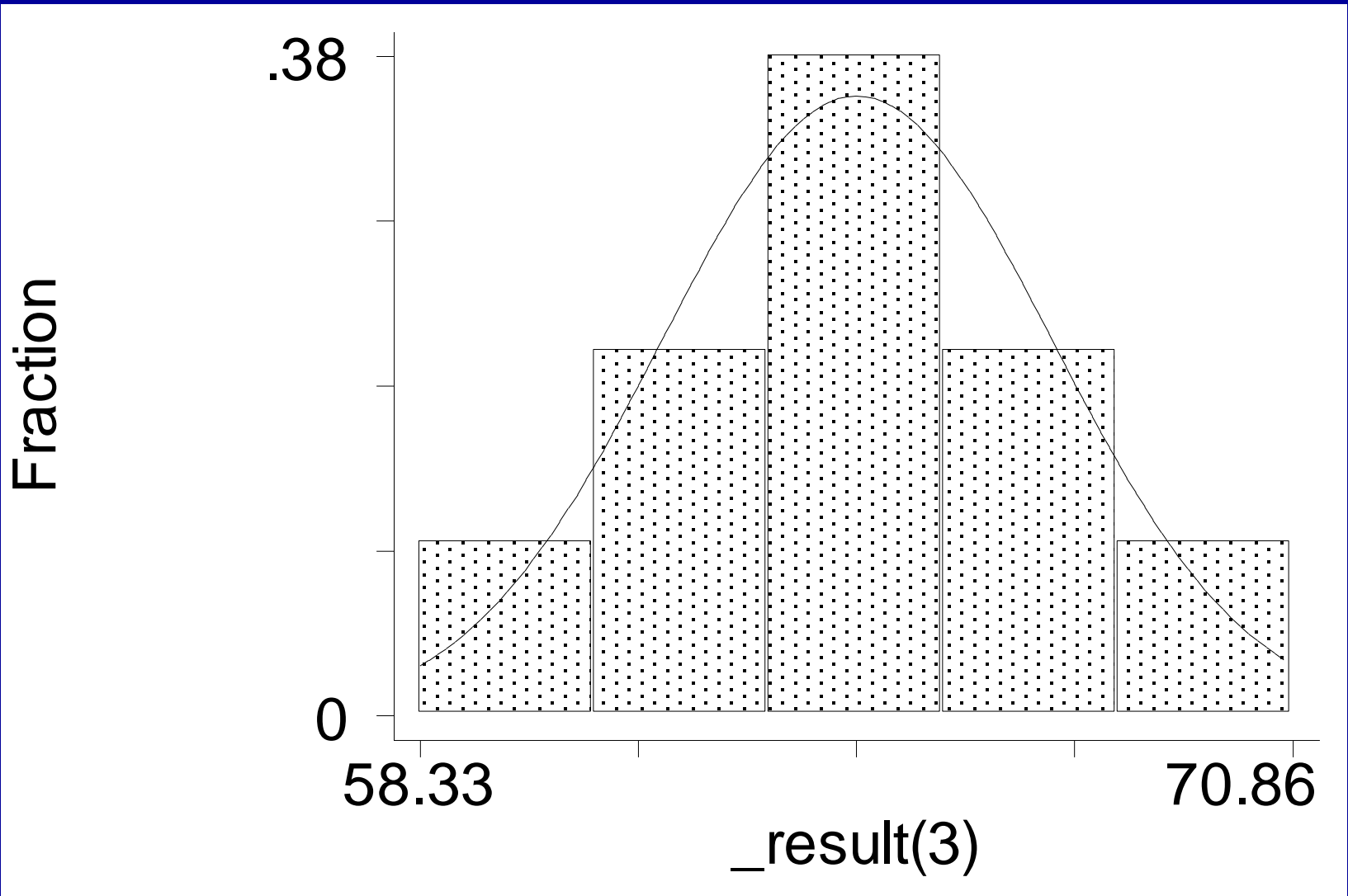


N = 2731; Mean = 64; sd = 29.6

Distribution of means: results of a sampling experiment (Mean – 64; sd – 29.6)

Size	Repeats	Min	Max	Mean	SD
100	100	58.3	70.8	64.6	2.8

Sample size – 100; Rpts – 100



Mean – 64.6; sd – 2.8

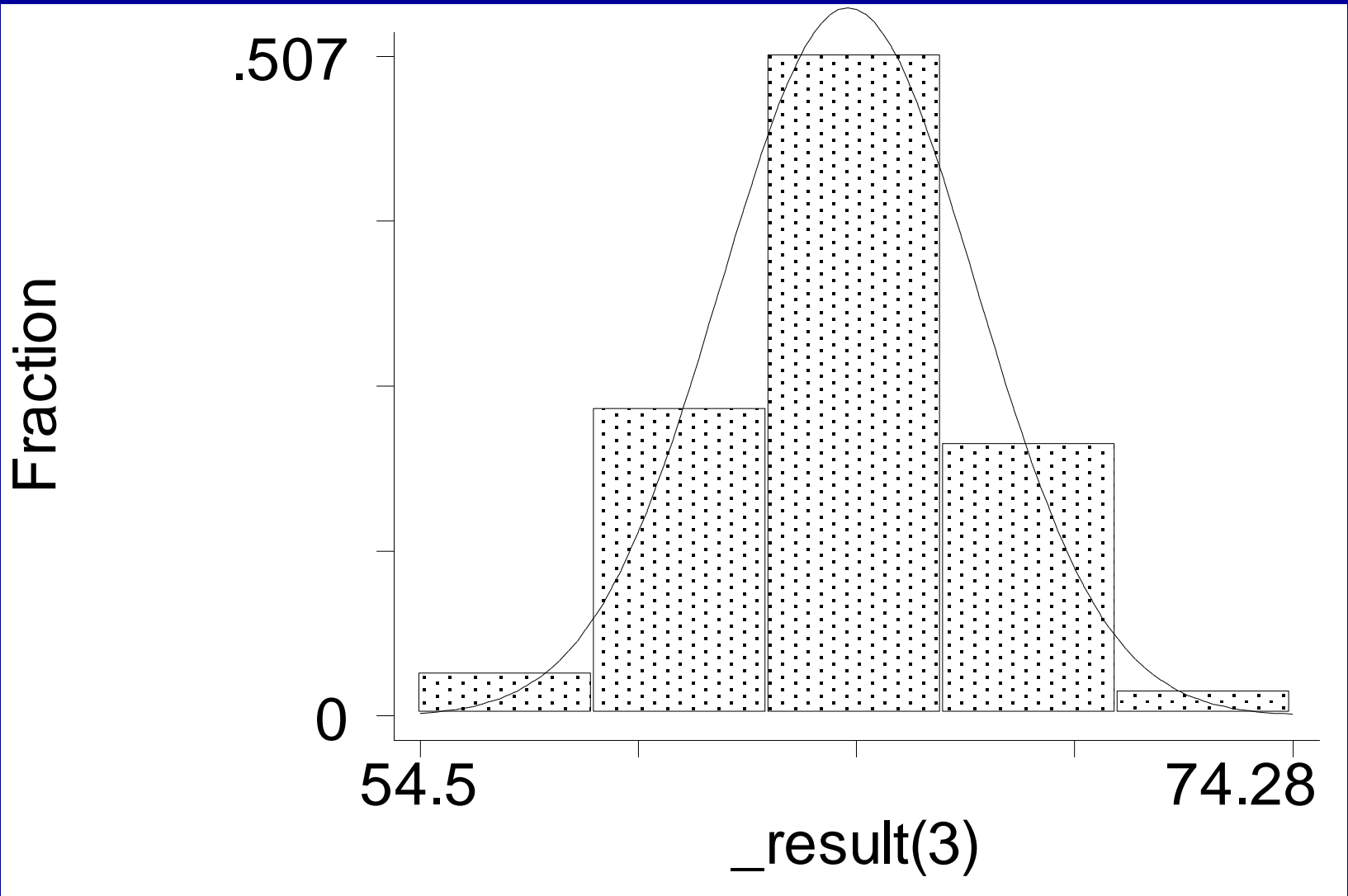
Distribution of means: results of a sampling experiment (Mean – 64; sd – 29.6)

Size	Repeats	Min	Max	Mean	SD
100	100	58.3	70.8	64.6	2.8
100	1000				

Distribution of means: results of a sampling experiment (Mean – 64; sd – 29.6)

Size	Repeats	Min	Max	Mean	SD
100	100	58.3	70.8	64.6	2.8
100	1000	54.5	74.3	64.2	2.9

Sample size – 100; Rpts - 1000



Mean – 64.2; sd – 2.9

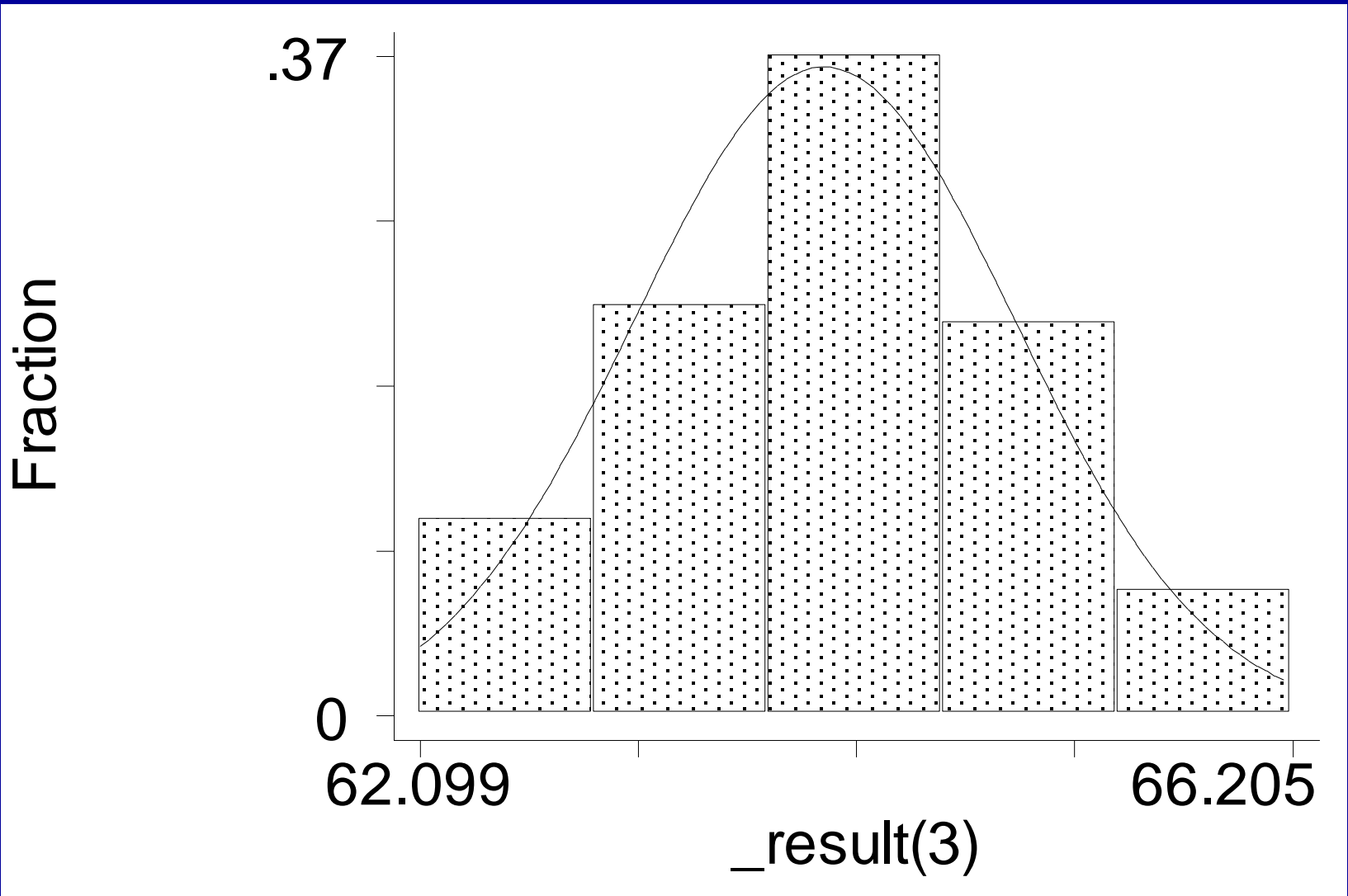
Distribution of means: results of a sampling experiment (Mean – 64; sd – 29.6)

Size	Repeats	Min	Max	Mean	SD
100	100	58.3	70.8	64.6	2.8
100	1000	54.5	74.3	64.2	2.9
1000	100				

Distribution of means: results of a sampling experiment (Mean – 64; sd – 29.6)

Size	Repeats	Min	Max	Mean	SD
100	100	58.3	70.8	64.6	2.8
100	1000	54.5	74.3	64.2	2.9
1000	100	62.1	66.2	64	0.9

Sample size – 1000; Rpts - 100



Mean – 64; sd – 0.9

- Sample means form a Normal distribution
- Mean of these means equal to the population mean
- The standard deviation of the sample means is called **standard error of mean**
- **Difference between two means, percentage & difference between two percentages**

Standard Error of Mean

(Standard deviation of sample means)

- Depends on

Variability – Standard Deviation

Sample size

Standard Error of the Mean OR Standard Deviation of Sample Means

- Repeated large samples of equal size taken to estimate a population mean will form a normal distribution with a standard error

$$SE_{mean} = \sqrt{\frac{SD^2}{n}}$$

Confidence Interval

- Based on the point estimate and standard error
- Limits which are likely to include population parameter (mean)
- Can be calculated for a mean, percentage, relative risk etc.
- Also for difference between two means or two percentages

A random sample of 100 infants born at Ragama hospital during 2003 had a mean birthweight of 2800 grams. The standard deviation was 200.

Based on this data what could we infer about the mean birth weight of all infants born at Ragama hospital during 2003?

- In this situation we can be *reasonably certain* that the value of the population mean will be *close* to the sample mean
- In this situation we can be *95% confident* that the *range of values between mean minus 2 SE to mean plus 2 SE* is likely to contain the population mean value.

Standard Error of a Percentage

- Repeated large samples of equal size taken to estimate a population percentage will form a normal distribution with a standard error

$$\sqrt{\frac{pq}{n}}$$

p – Percentage to be estimated

q – 100 – p

n – sample size

Sampling Methods

- Simple Random Sampling
- Systematic Sampling
- Cluster Sampling
- Stratified Sampling
- Multistage Sampling
- Area Sampling
- Quota Sampling
- Convenience Sampling



Simple Random Sampling

Requires a complete sampling frame (a listing of all units/elements) and random numbers.

Each element has equal probability of being selected.

Systematic Sampling

Random numbers are not required. Every n^{th} person from the sampling frame is selected.

Cluster sampling

The units are not selected in to the sample individually but as groups or clusters.

This may be done when a complete sampling frame is not available but it is known that the units belong to some larger groupings.

Provided -

Stratified sampling

When it is known that the characteristic under investigation is related to another variable, the population under study could be stratified according to the second variable.


Multistage sampling

Selecting the sample in more than one stage. Multistage sampling enables the investigator to concentrate/restrict the fieldwork to selected areas.

Sampling methods

- Area sampling
- Quota sampling
- Convenience sampling

- Snowball sampling
- Respondent driven sampling (RDS)

- All these are technical terms 
- In many instances combinations or modifications are necessary
- The sample should be representative of the population

1. Epsem sampling Vs Unequal probability of selection
2. Element sampling Vs Cluster sampling
3. Unstratified sampling Vs Stratified
4. Random selection Vs Systematic selection
5. Single stage Vs Multistage sampling

Sampling

- What is it?
- Why do it?
- How is it done?

Sampling Exercises

Describe how an appropriate sample should be selected for each given study.

What information is needed to calculate the required sample size?

1. Study to determine the alcohol consumption pattern of adults in a district
2. Study to determine prescription pattern in government hospital OPDs in a district
3. Study to determine nutritional status of Sri Lankan primary school children
4. Study to determine the quality of management of diabetes in clinics in the state sector
5. Study to map the distribution of Aedes breeding sites in the CMC area